Stat 544: Time Series Analysis
The University of Montana

Chapter 3: Multiple Linear Regression

Fall 2011

Why is Regression Important in Time Series?

- Polynomial Regression and Spline Regression are used to analyze trend (Chapter 4).
- Time Series Smoothing and Filtering are local regressions (Chapter 4)
- Autoregression and Transfer Function models (Chapters 6-8, 10).
- Frequency domain ideas follow from regression model–using sines and cosines as inputs. (Chapters 9 and 10)
- Extensions to filters of infinite extent can be handled using regression in the frequency domain.

The Multiple Linear Regression Model

- The multiple linear regression model is given by:

$$X_t = \beta_1 Z_{1t} + \beta_2 Z_{2t} + \cdots + \beta_q Z_{tq} + W_t$$

  where

$$\{W_t\} \sim IID(0, \sigma^2)$$

  for $t = 1, 2, \ldots, n$.
- There are $q + 1$ parameters ($\beta_1, \beta_2, \ldots, \beta_q$ and σ^2).
- Assumption in linear regression:
  1. Linearity in $\beta_1, \ldots, \beta_q$
  2. The covariates $Z_1, \ldots, Z_q$ are fixed (Partially unsatisfactory)
  3. Independence (Unsatisfactory)
  4. Homogeneity of variance over time.

The Multiple Linear Regression Model Cont'd...

- Transformations may help to alleviate the violations of assumptions (1) and (4).
- *A Special Case:* A simple linear regression,

$$X_t = \beta_1 + \beta_2 t + W_t.$$

Here $Z_{1t} = 1$ and $Z_{2t} = t$.

- Another example:

$$X_t = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 \cos(2\pi\omega t) + \beta_5 \sin(2\pi\omega t) + W_t$$

where ω is a known constant. What if it is not?

Example: Global Warming

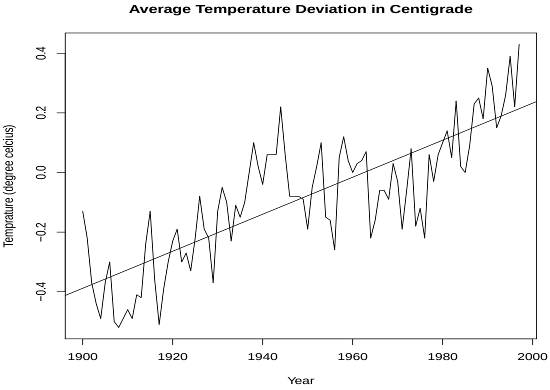


Figure: $Z_{1t} = 1$ and $Z_{2t} = t$

Example: Weekly Cardiovascular Mortality in LA ($n = 508$)

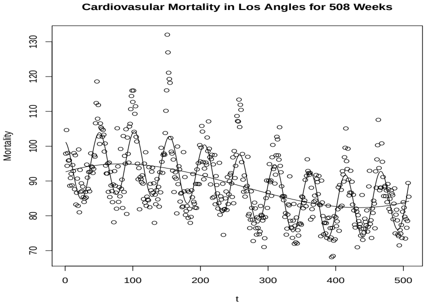


Figure: $Z_{1t} = 1$, $Z_{2t} = t$, $Z_{3t} = t^2$, $Z_{4t} = t^3$, $Z_{5t} = \cos\left(2\pi(\frac{1}{52})t\right)$ and $Z_{6t} = \sin\left(2\pi(\frac{1}{52})t\right)$

Recruitment versus Present and Past of SOI

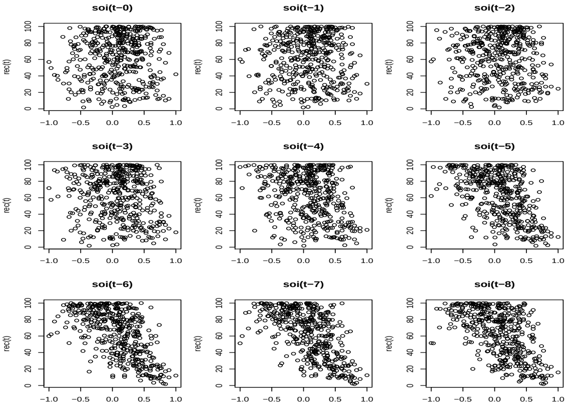


Figure: $Z_{1t} = 1$ and $Z_{2t} = \mathrm{SOI}_{t-6}$; R-code is lag.plot2(rec,soi,8)

The Multiple Linear Regression Model in Vector Form

- The multiple regression model can be written in a vector form as,

$$X_t = \boldsymbol{\beta}' \mathbf{Z}_t + W_t$$

where

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{pmatrix} \quad \text{and} \quad \mathbf{Z}_t = \begin{pmatrix} Z_{t1} \\ Z_{t2} \\ \vdots \\ Z_{tq} \end{pmatrix}.$$

Ordinary Least Square

- Define the matrices,

$$\mathbf{Z} = \begin{pmatrix} Z_{11} & Z_{12} & Z_{13} & \cdots & Z_{1q} \\ Z_{21} & Z_{22} & Z_{23} & \cdots & Z_{2q} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ Z_{n1} & Z_{n2} & Z_{n3} & \cdots & Z_{nq} \end{pmatrix} \quad \text{and} \quad \mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}.$$

- The estimator for the coefficients vector is,

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_q \end{pmatrix} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}.$$

- This estimator is **always** unbiased, i.e. $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$.

Ordinary Least Square Cont'd...

- The covariance of $\hat{\boldsymbol{\beta}}$,
$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma_W^2(\mathbf{Z}'\mathbf{Z})^{-1}.$$
- We need the assumption the $\{W_t\} \sim WN(0, \sigma_W^2)$ for the above formula to hold.
- Therefore, the errors are correlated or heteroscedastic, the estimated covariance of the parameters will be incorrect. Methods to deal with this will be covered in Chapter 6 and 7.
- Define the residual sum of squares RSS as
$$RSS = (\mathbf{X} - \mathbf{Z}\hat{\beta})'(\mathbf{X} - \mathbf{Z}\hat{\beta}) = \mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}.$$
- An unbiased estimator of σ_W^2, when $\{W_t\} \sim WN(0, \sigma_W^2)$, is
$$s^2 = \frac{RSS}{n-q}.$$

Tests on Individual Parameters

- In this section, we assume that $\{W_t\}$ is a Gaussian White Noise, i.e. $W_t \sim N(0, \sigma_W^2)$ for all t.
- To test the hypothesis: $H_0 : \beta_i = \beta_i^{(0)}$:

$$t_i = \frac{\hat{\beta}_i - \beta_i^{(0)}}{s\sqrt{c_{ii}}} \overset{H_0}{\sim} t_{n-q}$$

where c_{ii} is the ith diagonal element of $C = (X'X)^{-1}$.

- If the errors are correlated, $s\sqrt{c_{ii}}$ needs to be adjusted and, yet, inference procedures will be suboptimal.

Example: Global Temperature

```
> t<-1:length(gtemp)
> fit_1<-lm(gtemp~t)
> summary(fit_1)

Call: lm(formula = gtemp ~ t)

Residuals:
       Min         1Q     Median         3Q        Max
-0.3194611 -0.0972175  0.0008361  0.0824544  0.2938335

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.3981395  0.0220798  -18.03   <2e-16 ***
         t   0.0057485  0.0002925   19.65   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.1251 on 128 degrees of freedom Multiple
R-squared: 0.7511,     Adjusted R-squared: 0.7492 F-statistic: 386.3
on 1 and 128 DF,  p-value: < 2.2e-16
```

Example: Weekly Cardiovascular Mortality in LA

```
> t <- 1:length(cmort)
> c_t=cos(2*pi*(1/52)*t)
> s_t=sin(2*pi*(1/52)*t)
> t2<-t^2
> t3<-t^3
> fit_cmort<-lm(cmort~t+t2+t3+c_t+s_t)
> summary(fit_cmort)

Call: lm(formula = cmort ~ t + t2 + t3 + c_t + s_t)

...

Coefficients:
              Estimate  Std. Error  t value   Pr(>|t|)
(Intercept)  9.242e+01  1.088e+00  84.947    < 2e-16 ***
 t           7.204e-02  1.849e-02   3.895    0.000111 ***
 t2         -5.184e-04  8.438e-05  -6.144    1.64e-09 ***
 t3          6.943e-07  1.090e-07   6.371    4.26e-10 ***
 c_t         8.901e+00  3.821e-01  23.295    < 2e-16 ***
 s_t        -1.803e+00  3.811e-01  -4.730    2.92e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 6.052 on 502 degrees of freedom Multiple
R-squared: 0.6373,     Adjusted R-squared: 0.6337 F-statistic: 176.4
on 5 and 502 DF,  p-value: < 2.2e-16
```

Testing Significance of a Group of Parameters

- Assume $\{W_t\} \sim GWN(0, \sigma_W^2)$.
- We want to test $H_0 : \beta_{q_1+1} = \cdots = \beta_q = 0$ for some $q_1 < q$.
- The reduced model under the hypothesis is

$$X_t = \beta_1 Z_{t1} + \beta_2 Z_{t2} + \cdots + \beta_{q_1} Z_{tq_1} + W_t.$$

- Let RSS_1 be the residual sum of squares for the reduced model. (When we fit the reduced model there will be an increase in the residual sum of square i.e $RSS_1 > RSS$.)
- The significance of the q_1 variables may be tested by

$$F = \frac{RSS_1 - RSS}{RSS} \frac{n-q}{q-q_1} \sim F_{q-q_1, n-q}.$$

- The quantity $RSS_1 - RSS$ is called *partial sum of squares*.

Summarizing Results: ANOVA Table

Source	df	Sum of Squares	Mean Squares	F
$Z_{t,q_1+1}, \ldots Z_{tq}$	$q - q_1$	$RSS_1 - RSS$	$(RSS_1 - RSS)/(q - q_1)$	?
Error	$n - q$	RSS	$s^2 = RSS/(n - q)$	
Total	$n - q_1$	RSS_1		

Testing Significance of a Group of Parameters: Example

- Consider the quarterly per share earnings of Johnson& Johnson.
- For $i = 2, 3, 4$, let $Q_i = 1$ for the ith quarter and 0 otherwise.

$$\log(X_t) = \beta_1 + \beta_2 t + \beta_3 Q_2 + \beta_4 Q_3 + \beta_5 Q_4 + W_t.$$

- To test for the significance of the seasonal effects,

$$H_0 : \beta_3 = \beta_4 = \beta_5 = 0.$$

- To test for the significance of the global trend,

$$H_0 : \beta_2 = 0.$$

Example: Per Share Quarterly Earning J & J

```
> JJ<-log(scan("jj.dat"))
Read 84 items
> t<-1:length(JJ)
> Q_2<-rep(c(0,1,0,0),21)
> Q_3<-rep(c(0,0,1,0),21)
> Q_4<-rep(c(0,0,0,1),21)
>
> fit_JJ_full<-lm(JJ~t+Q_2+Q_3+Q_4)
> fit_JJ_red<-lm(JJ~t)
> anova(fit_JJ_red,fit_JJ_full)
Analysis of Variance Table

Model 1: JJ ~ t
Model 2: JJ ~ t + Q_2 + Q_3 + Q_4
     Res.Df     RSS    Df   Sum of Sq      F    Pr(>F)
  1     82    2.06104
  2     79    1.24181  3   0.81923       17.372 9.357e-09 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Testing Significance of a Group of Parameters

- A special case of interest is when the reduced model is

$$X_t = \beta_1 + W_t.$$

  That is, $q_1 = 1$ and $Z_{1t} = 1$.
- Let the residual sum of squares under the reduced model be denoted by RSS_0.
- It can be shown that $RSS_0 = \sum_{t=1}^{n} (X_t - \bar{X})^2$.
- The quantity

$$R^2 = \frac{RSS_0 - RSS}{RSS_0}$$

  measures the proportion of variation explained by the regression.

Sequential Methods

- Goal: To identify a subset of useful variable in explaining and predicting the dependent variable.
- Methods that proceed sequentially:
  - Forward Selection
  - Backward elimination
  - Stepwise regression
- Variables are added or deleted when the p-value of the variable from the F-test either exceeds or fails to exceed some predetermined levels.
- Basically nested models are compared.
- The hypothesis testing in the previous section can be used to test whether the drop in fit when a group of variables are deleted is significant.

Non–Sequential Methods

- Evaluate each of the candidate models on its own merits.
- The models to be compared do not have to be nested as long as they are based on the same sample.
- Consider a regression model with k-coefficients and denote

$$\hat{\sigma}_k^2 = \frac{RSS_k}{n}$$

  is another estimate of σ_W^2 known as the Maximum Likelihood estimate (MLE).
- $\hat{\sigma}_k^2$ can be used to compare models but one major drawback it has is that it monotonically decreases as k increases.

Model Selection Criteria: AIC and AICC

- *Akaike Information Criterion (AIC)* is defined as

$$AIC = \log \hat{\sigma}_k^2 + \frac{n + 2k}{n}.$$

- The second term in the right hand side is called *penalty term*.
- *Bias Corrected AIC*

$$AICC = \log \hat{\sigma}_k^2 + \frac{n + k}{n - k - 2}$$

- The smaller is the better.

Model Selection Criteria:SIC and Others

- *Schwarz's Information Criterion (SIC)*

$$SIC = \log \hat{\sigma}_k^2 + \frac{k \log n}{n}$$

- *SIC* is also called *Bayesian Information Criterion (BIC).*
- Empirical studies show that AIC is preferable for smaller samples with large # of parameters and SIC for large sample .
- Other model selection criteria include Adjusted-R^2 and Mallows' C_p

Cardiovascular Mortality in Los Angeles.

A study to see the possible effects of temperature and particulate pollution on the daily cardiovascular mortality on Los Angeles.

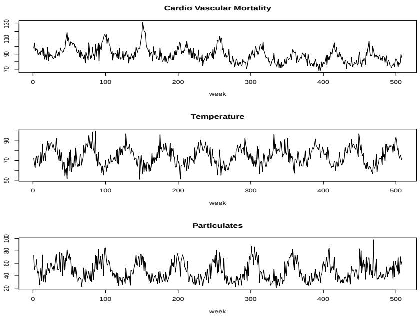
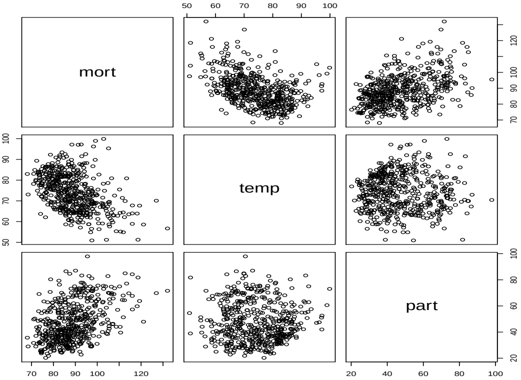


Figure: Time Plot for Particulate, Temperature and Mortality

Scatter Plot Matrix for Particulate, Temperature and Mortality



Model Specification

- There is a strong seasonal component for all the series corresponding to winter-summer variation.
- There is a downward trend in the cardiovascular mortality data.
- Based on the time plot and scatter plot matrix we entertain the following candidate models:

$$M_t = \beta_o + \beta_1 t + W_t \tag{1}$$

$$M_t = \beta_0 + \beta_1 t + \beta_2(T_t - \bar{T}) + W_t \tag{2}$$

$$M_t = \beta_0 + \beta_1 t + \beta_2(T_t - \bar{T}) + \beta_3(T_t - \bar{T})^2 + W_t \tag{3}$$

$$M_t = \beta_0 + \beta_1 t + \beta_2(T_t - \bar{T}) + \beta_3(T_t - \bar{T})^2 + \beta_4 P_t + W_t \tag{4}$$

Model Selection

- Model fit summary

Table: Summary Statistics for Mortality Models

Model	RSS	s^2	R^2	AICC
1	40,020	79.09	0.21	5.38
2	31,413	62.20	0.38	5.14
3	27,985	55.52	0.45	5.03
4	20,509	40.77	0.60	4.72

- 60% explained variation for model 4.

Model Comparisons

- The best model (according to s^2, R^2 and $AICC$)

$$\hat{M}_t = 81.59 - 0.027_{(0.002)} t - 0.473_{(0.032)}(T_t - 74.6) + 0.023_{(0.003)}(T_t - 74.6)^2 + 0.255_{(0.019)} P_t.$$

- How do we interpret estimates of the parameter and do they concur with the observations from the scatter plot matrix and the time plot?
- For example, a 10 degrees increase in temperature will be associated with about 5 deaths daily, other variables remaining constant.
- The residuals $\hat{w}_t = M_t - \hat{M}_t$ can be checked for autocorrelation.

Model Comparison: Nested Models

- One can compare nested models using the residual sum of squares.
- To compare the model with only trend against the full model. Here, $q = 5$, $q_1 = 2$ and $n = 508$. Then

$$F_{3,503} = \frac{40,020 - 20,509}{20,509}\frac{503}{3} = 160$$

- $F_{3,\infty}(0.001) = 5.42$. What is the conclusion?

R-Script

```
mort=scan("cmort.dat")
temp=scan("temp.dat")
part=scan("part.dat")
par(mfrow=c(3,1))
ts.plot(mort, main="Cardio Vascular Mortality",xlab="week",ylab="")
ts.plot(temp,main="Temperature",xlab="week",ylab="")
ts.plot(part, main="Particulates",xlab="week",ylab="")
par(mfrow=c(1,1))
pairs(cbind(mort,temp,part)) # scatter plot matrix
temp=temp-mean(temp)
temp2=temp^2
t=1:length(mort)
fit=lm(mort t+temp+temp2+part)
summary(fit)
AIC(fit)/508 #R gives n*AIC
```

Average August Discharge of a River in Many Glacier, MT

- The objective is to see if Air Temperature and Annual Precipitation predict the August Discharge of the river.
- Data is available for 59 Years.
- Discharge is measures in Cubic Feet per Second (CFC)
- Air Temperature is measured in Degrees Fahrenheit (dF)
- Precipitation is measure in Inches (in).
- The data was sent to you by email.

Homework:

1. For your own data from chapter 1, apply the methods of this Chapter.
2. Problems 2.1-2.3 and Problem 2.11 (a), (b), (c), (e) and (f). For part (e) ignore the nonparametric smoothing part.